ELSEVIER

Journal of Computational and Applied Mathematics 138 (2002) 127–150

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

A primal–dual algorithm for minimizing a sum of Euclidean norms ☆

Liqun Qi[a,*], Defeng Sun[b], Guanglu Zhou[b]

[a] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China*

[b] *Department of Mathematics, National University of Singapore, Singapore*

Received 28 February 2000; received in revised form 24 November 2000

Abstract

We study the problem of minimizing a sum of Euclidean norms. This nonsmooth optimization problem arises in many different kinds of modern scientific applications. In this paper we first transform this problem and its dual problem into a system of strongly semismooth equations, and give some uniqueness theorems for this problem. We then present a primal–dual algorithm for this problem by solving this system of strongly semismooth equations. Preliminary numerical results are reported, which show that this primal–dual algorithm is very promising. © 2002 Elsevier Science B.V. All rights reserved.

MSC: 90C33; 90C30; 65H10

Keywords: Primal–dual algorithm; Sum of norms; Semismooth; VLSL design; Euclidean facilities location; Steiner minimum trees

1. Introduction

Consider the problem of minimizing a sum of Euclidean norms:

$$\min_{x\in\Re^n} \sum_{i=1}^{m} ||b_i - A_i^{\mathrm{T}}x||, \tag{1.1}$$

where $b_1, b_2, \ldots, b_m \in \Re^d$ are column vectors in the Euclidean d-space, $A_1, A_2, \ldots, A_m \in \Re^{n\times d}$ are n-by-d matrices. Let $A = [A_1, A_2, \ldots, A_m]$ and $b^{\mathrm{T}} = [b_1^{\mathrm{T}}, \ldots, b_m^{\mathrm{T}}]$. Let

$$f(x) = \sum_{i=1}^{m} f_i(x), \tag{1.2}$$

☆ This work is partially supported by the Australian Research Council and the Research Grant Council of Hong Kong.

* Corresponding author.

E-mail addresses: maqilq@polyu.edu.hk (L. Qi), matsundf@nus.edu.sg (D. Sun), smazgl@nus.edu.sg (G. Zhou).

0377-0427/02/$ - see front matter

PII: S0377-0427(01)00357-0

where

$$f_i(x) = ||b_i - A_i^{\mathrm{T}}x||, \quad \text{for } i = 1, \ldots, m. \tag{1.3}$$

It is clear that $x=0$ is an optimal solution to problem (1.1) when all of the b_i are zero. Therefore, we assume in the rest of this paper that not all of the b_i are zero. Problem (1.1) is a convex programming problem, but its objective function f is not differentiable at any point x where $b_i - A_i^{\mathrm{T}}x = 0$ for some i. Problem (1.1) arises in many applications, such as the VLSL design [1], the Euclidean facilities location problem and the Steiner minimal tree problem under a given topology [29]. Many algorithms have been designed to solve problem (1.1), see [1–5,18,26,28–30].

In this paper we study the problem of minimizing a sum of norms. First in Section 2 we transform this problem and its dual problem into some strongly semismooth equations. These transformations are very important to our design of quadratically convergent algorithms. In Section 3 we give some conditions for this problem having a unique solution. These conditions will be applied to the quadratic convergence analysis of the algorithm presented in Section 5. In Section 4 we present an augmented smoothing algorithm for solving nonsmooth equations, which is an extension of the smoothing methods proposed in [19,24,27,32]. In particular, global and superlinear convergence of this algorithm is established under some weaker conditions than those in [24]. In Section 5 we then propose a primal–dual algorithm for minimizing a sum of Euclidean norms by applying the augmented smoothing algorithm presented in Section 4 directly to a system of strongly semismooth equations derived from this problem and its dual problem. In Section 6, some numerical results are reported. We conclude this paper in Section 7.

Some words about our notation. For a continuously differentiable function $F:\Re^n \to \Re^m$, we denote the Jacobian of F at $x \in \Re^n$ by $F'(x)$, whereas the transposed Jacobian as $\nabla F(x)$. In particular, if $m=1$, the gradient $\nabla F(x)$ is viewed as a column vector.

Let $F:\Re^n \to \Re^m$ be a locally Lipschitzian vector function. By Rademacher's theorem, F is differentiable almost everywhere. Let Ω_F denote the set of points where F is differentiable. Then the B-subdifferential of F at $x \in \Re^n$ is defined to be

$$\partial_B F(x) = \left\{ \lim_{\substack{x^k \to x \\ x^k \in \Omega_F}} \nabla F(x^k)^T \right\}, \tag{1.4}$$

while Clarke's generalized Jacobian of F at x is defined to be

$$\partial F(x) = \operatorname{conv} \partial_B F(x), \tag{1.5}$$

(see [11,21,25]). F is called *semismooth* at x if F is directionally differentiable at x and for all $V \in \partial F(x+h)$ and $h \to 0$,

$$F'(x;h) = Vh + o(||h||); \tag{1.6}$$

F is called *p-order semismooth*, $p \in (0,1]$, at x if F is semismooth at x and for all $V \in \partial F(x+h)$ and $h \to 0$,

$$F'(x;h) = Vh + O(||h||^{1+p}); \tag{1.7}$$

F is called *strongly semismooth* at x if F is 1-order semismooth at x. F is called a (strongly) semismooth function if it is (strongly) semismooth everywhere (see [21,25]). Here, $o(||h||)$ stands for a vector function $e:\mathfrak{R}^n \to \mathfrak{R}^m$, satisfying

$$\lim_{h\to 0} \frac{e(h)}{||h||} = 0,$$

while $O(||h||^2)$ stands for a vector function $e:\mathfrak{R}^n \to \mathfrak{R}^m$, satisfying

$$||e(h)|| \leqslant M||h||^2$$

for all h satisfying $||h|| \leqslant \delta$, and some $M > 0$ and $\delta > 0$.

Lemma 1.1 (Qi and Sun [25]). (i) *If F is semismooth at x, then for any $V \in \partial F(x+h)$ and any $h \to 0$,*

$$F(x+h) - F(x) - Vh = o(||h||);$$

(ii) *If F is p-order semismooth at x, then for any $V \in \partial F(x+h)$ and any $h \to 0$,*

$$F(x+h) - F(x) - Vh = O(||h||^{1+p}).$$

For a convex set $\Omega \subset \mathfrak{R}^n$, $\Pi_\Omega(\cdot)$ is the projection operator onto Ω. For a vector $x \in \mathfrak{R}^n$, $||x||$ represents the Euclidean norm $(\sum_{i=1}^n x_i^2)^{1/2}$. Let I_d denote the $d \times d$ identity matrix. Let $\mathfrak{R}_+ = \{\epsilon \in \mathfrak{R}\colon \epsilon \geqslant 0\}$ and $\mathfrak{R}_{++} = \{\epsilon \in \mathfrak{R}\colon \epsilon > 0\}$. Finally, we use $\epsilon \downarrow 0^+$ to denote the case that a positive scalar ϵ tends to 0.

2. Some equivalent formulations

The dual of problem (1.1) has the following form:

$$\max_{y\in Y} b^{\mathrm{T}} y, \tag{2.1}$$

where

$$Y = \{y = [y_1^{\mathrm{T}}, \ldots, y_m^{\mathrm{T}}]^{\mathrm{T}} \in \mathfrak{R}^{md}\colon\ y_i \in \mathfrak{R}^d, \quad ||y_i|| \leqslant 1,\ i = 1, \ldots, m; Ay = 0\}. \tag{2.2}$$

Theorem 2.1 (see Andersen [2]). *Let $x \in \mathfrak{R}^n$, $y \in Y$ and let $x^* \in \mathfrak{R}^n$, $y^* \in Y$ be optimal solutions to problems* (1.1) *and* (2.1), *respectively. Then*

$$(a) \quad b^{\mathrm{T}} y \leqslant \sum_{i=1}^m ||b_i - A_i^{\mathrm{T}} x|| \quad \text{(weak duality)}$$

and

$$(b) \quad b^{\mathrm{T}} y^* = \sum_{i=1}^m ||b_i - A_i^{\mathrm{T}} x^*|| \quad \text{(strong duality).}$$

Definition 2.1 (see Andersen [2]). A point $x \in \Re^n$ and a point $y \in Y$ are called ϵ-optimal to problems (1.1) and (2.1) if

$$\sum_{i=1}^{m} ||b_i - A_i^{\mathrm{T}}x|| - b^{\mathrm{T}}y \leqslant \epsilon.$$

From Theorem 2.1 we have that (x^*, y^*) is a pair of optimal solutions to problems (1.1) and (2.1) if and only if (x^*, y^*) is a solution to the following equations and inequations:

$$\begin{aligned} &Ay = 0, \\ &||y_i|| \leqslant 1, \quad i = 1, \dots, m, \\ &\sum_{i=1}^{m} ||b_i - A_i^{\mathrm{T}}x|| - b^{\mathrm{T}}y = 0. \end{aligned} \tag{2.3}$$

In [4] it was proved that (2.3) is equivalent to

$$\begin{aligned} &Ay = 0, \\ &||y_i|| \leqslant 1, \quad i = 1, \dots, m, \\ &(b_i - A_i^{\mathrm{T}}x) - ||b_i - A_i^{\mathrm{T}}x|| y_i = 0, \quad i = 1, \dots, m. \end{aligned} \tag{2.4}$$

It follows from (2.4) that if (x^*, y^*) is a pair of optimal solutions to problems (1.1) and (2.1), then for $i = 1, \dots, m$, either $b_i - A_i^{\mathrm{T}}x^* = 0$ or $||y_i^*|| = 1$. If $b_i - A_i^{\mathrm{T}}x^* \neq 0$ for some i then $y_i^* = \nabla f_i(x^*)$.

Definition 2.2. Let (x^*, y^*) be a pair of optimal solutions to problems (1.1) and (2.1). We say strict complementarity holds at (x^*, y^*) if, for each i, $||y_i^*|| < 1$ when $b_i - A_i^{\mathrm{T}}x^* = 0$.

In [26] it was proved that (2.3) is equivalent to

$$\begin{aligned} &Ay = 0, \\ &y_i - \Pi_B(y_i + (b_i - A_i^{\mathrm{T}}x)) = 0, \quad i = 1, \dots, m, \end{aligned} \tag{2.5}$$

where $B = \{s \in \Re^d \colon ||s|| \leqslant 1\}$ and Π_B is the projection operator onto B.

Lemma 2.1. *Let $\Omega \subset \Re^n$ be a convex set. If $s, t \in \Re^n$ satisfy $s = \Pi_\Omega(s + t)$, then $r = s + t$ and t satisfy $t - r + \Pi_\Omega(r) = 0$. Conversely, if r and t satisfy $t - r + \Pi_\Omega(r) = 0$ then $s = \Pi_\Omega(r)$ and t satisfy $s = \Pi_\Omega(s + t)$.*

Proof. If $s, t \in \Re^n$ satisfy $s = \Pi_\Omega(s + t)$, then

$$\Pi_\Omega(r) = \Pi_\Omega(s + t) = s = r - t.$$

So, $t - r + \Pi_\Omega(r) = 0$. Conversely, if r and t satisfy $t - r + \Pi_\Omega(r) = 0$, then

$$\Pi_\Omega(s + t) = \Pi_\Omega(\Pi_\Omega(r) + t) = \Pi_\Omega(r - t + t) = \Pi_\Omega(r) = s. \qquad \square$$

Let $P(y) = [\Pi_B(y_1)^{\mathrm{T}}, \Pi_B(y_2)^{\mathrm{T}}, \dots, \Pi_B(y_m)^{\mathrm{T}}]^{\mathrm{T}}$. It follows from Lemma 2.1 that (2.5) is equivalent to

$$\begin{aligned} &AP(y) = 0, \\ &y - P(y) - (b - A^{\mathrm{T}}x) = 0, \end{aligned} \tag{2.6}$$

in the sense that if (x^*, y^*) is a solution of (2.6) then $(x^*, P(y^*))$ is a solution of (2.5), and conversely if (x^*, y^*) is a solution of (2.5) then $(x^*, y^* + (b - A^T x^*))$ is a solution of (2.6). Let

$$F(x,y) = \begin{pmatrix} AP(y) \\ y - P(y) - (b - A^T x) \end{pmatrix}. \tag{2.7}$$

So, we can solve problems (1.1) and (2.1) by solving the following nonsmooth equations:

$$F(x,y) = 0. \tag{2.8}$$

3. Uniqueness theorems

In what follows we always assume that A has rank n. Under this assumption, the solution set of problem (1.1) is nonempty and bounded; see Lemma 2.1 [26]. However, it is easy to show that A having rank n is not sufficient for problem (1.1) having a unique solution. For example, let $x \in \Re^2$, $b_1 = (0,1)^T$, $b_2 = (0,-1)^T$, $A_1 = A_2 = I_2$, $A = [A_1, A_2]$, and

$$f(x) = ||b_1 - A_1^T x|| + ||b_2 - A_2^T x||.$$

Clearly, A has rank 2, and $f(x)$ has a minimum of 2 which is attained for all $x = (0, x_2)^T$ with $-1 \leqslant x_2 \leqslant 1$. In this section we will give some conditions for problem (1.1) having a unique solution.

Lemma 3.1. (i) *For any* $x \in \Re^n$ *and* $y \in \Re^n$ *with* $||y|| \leqslant 1$, $||x|| \geqslant y^T x$. (ii) *If* $x \neq 0$, *then* $||x|| > y^T x$ *for any* $y \in \Re^n$ *with* $||y|| < 1$.

Proof. Since $||x|| = \max_{||y||=1} y^T x = \max_{||y|| \leqslant 1} y^T x$, (i) and (ii) hold. □

Theorem 3.1. *Let* (x^*, y^*) *be a pair of optimal solutions to problems* (1.1) *and* (2.1). x^* *is a unique solution to problem* (1.1) *if*

(i) *strict complementarity holds at* (x^*, y^*), *and*

(ii) $\bar{A} = [A_i,\ i \in M_0(x^*)]$ *has rank* n, *where*

$$M_0(x^*) = \{i\colon ||b_i - A_i^T x^*|| = 0,\ i = 1, \ldots, m\}.$$

Proof. For any $d \in \Re^n$ with $d \neq 0$ and $\delta > 0$ small enough,

$$\begin{aligned} f(x^* + \delta d) &= \sum_{i=1}^m ||b_i - A_i^T (x^* + \delta d)|| \\ &\geqslant f(x^*) + \sum_{i \in M_0(x^*)} \delta ||A_i^T d|| - \sum_{i \notin M_0(x^*)} \delta (y_i^*)^T (A_i^T d) \\ &= f(x^*) + \sum_{i \in M_0(x^*)} \delta ||A_i^T d|| - \sum_{i \notin M_0(x^*)} \delta (A_i y_i^*)^T d \\ &= f(x^*) + \delta \sum_{i \in M_0(x^*)} \{||A_i^T d|| + (y_i^*)^T (A_i^T d)\}, \quad \text{by (2.3)} \\ &> f(x^*), \quad \text{by (i), (ii) and Lemma 3.1 (ii).} \end{aligned} \tag{3.1}$$

The first inequality follows from the convexity of $f_i(x)$. (3.1) shows that x^* is a unique solution of problem (1.1). □

Theorem 3.2. *Let* (x^*, y^*) *be a pair of optimal solutions to problems* (1.1) *and* (2.1). *Suppose that* $\bar{A} = [A_i,\ i \in M_0(x^*)]$ *is an* $n \times n$ *nonsingular matrix, where*

$$M_0(x^*) = \{i\colon ||b_i - A_i^{\mathrm{T}} x^*|| = 0,\ i = 1, \ldots, m\}.$$

Then x^* *is a unique solution to problem* (1.1) *if and only if* $||y_i^*|| < 1$, *for all* $i \in M_0(x^*)$.

Proof. From Theorem 3.1, if $||y_i^*|| < 1$, for all $i \in M_0(x^*)$, then x^* is a unique solution to problem (1.1). On the other hand, if there exists $i \in M_0(x^*)$ such that $||y_i^*|| = 1$, then a unique $d \in \Re^n$ and $d \neq 0$ can be found satisfying

$$A_i^{\mathrm{T}} d = -y_i^*,$$
$$A_i^{\mathrm{T}} d = 0, \quad i \in M_0(x^*)\backslash\{i\},$$

and this corresponds to a direction of nonuniqueness as then it gives equality in (3.1) for all $\delta > 0$ small enough. □

Remark. When $d = 1$, problem (1.1) is the following linear ℓ_1 problem, see [17].

$$\min_{x \in R^n} \sum_{i=1}^{m} |b_i - a_i^{\mathrm{T}} x|, \tag{3.2}$$

where $b_1, b_2, \ldots, b_m \in \Re$, and $a_1, a_2, \ldots, a_m \in \Re^n$. Let $A = [a_1, a_2, \ldots, a_m]$ and $b^{\mathrm{T}} = [b_1, b_2, \ldots, b_m]$. From (2.1) the dual of problem (3.2) has the following form:

$$\max_{y \in Y} b^{\mathrm{T}} y, \tag{3.3}$$

where

$$Y = \{y \in \Re^m\colon |y_i| \leqslant 1,\ i = 1, \ldots, m;\ Ay = 0\}. \tag{3.4}$$

It follows from Theorem 3.2 that

Corollary 3.1 (Osborne [17]). *Let* (x^*, y^*) *be a pair of optimal solutions to problems* (3.2) *and* (3.3). *Suppose that* $\bar{A} = [a_i,\ i \in M_0(x^*)]$ *is an* $n \times n$ *nonsingular matrix, where*

$$M_0(x^*) = \{i\colon |b_i - a_i^{\mathrm{T}} x^*| = 0,\ i = 1, \ldots, m\}.$$

Then x^* *is a unique solution to problem* (3.2) *if and only if* $|y_i^*| < 1$, *for all* $i \in M_0(x^*)$.

4. An augmented smoothing algorithm for nonsmooth equations

In this section we present an augmented smoothing algorithm for solving nonsmooth equations, which is an extension of the smoothing methods proposed in [19,24,27,32]. In particular, global and superlinear convergence of this algorithm is established under some weaker conditions than that in [24].

Throughout this section, we assume that $\Phi:\Re^n \to \Re^n$ is a locally Lipschitzian function. We are interested in finding zeroes of Φ, i.e., solving the system of nonlinear equations

$$\Phi(x) = 0. \tag{4.1}$$

Such a system of nonlinear equations arises from nonlinear complementarity, variational inequality, nonlinear programming, the maximal monotone operation problem, nonsmooth partial differential equations, the nonsmooth compact fixed point problem, and the Newton method for the complex eigenvalue problem, the problem of minimizing a sum of norms; see [6–10,12–16,19–27,31,32] Various nonsmooth variants of Newton's methods, quasi-Newton methods, and Gauss–Newton methods, have been proposed and studied; see [21–23,25].

Definition 4.3. Let $z = (t,x) \in \Re \times \Re^n$. We say $H : \Re^{n+1} \to \Re^{n+1}$, defined by

$$H(z) := \begin{pmatrix} t \\ G(z) \end{pmatrix}, \tag{4.2}$$

an augmented smoothing function of Φ if

(i) G is a locally Lipschitzian function and G is continuously differentiable at any $(t,x) \in \Re^{n+1}$ with $t \neq 0$;
(ii) $G(0,x) = \Phi(x)$ for any $x \in \Re^n$.

Clearly, if H is an augmented smoothing function of Φ, then H is a locally Lipschitzian function and H is continuously differentiable at any $(t,x) \in \Re^{n+1}$ with $t \neq 0$. Moreover, $H(t^*,x^*) = 0$ if and only if $t^* = 0$ and $\Phi(x^*) = 0$. So solving the system of nonlinear equations (4.1) is equivalent to finding the zeros of H.

Choose $\bar{t} \in \Re_{++}$ and $\gamma \in (0,1)$ such that $\gamma\bar{t} < 1$. Let $\bar{z} := (\bar{t}, 0) \in \Re \times \Re^n$ and $s \in [\frac{1}{2}, 1]$. Define the merit function $\psi : \Re^{n+1} \to \Re_+$ by

$$\psi(z) := ||H(z)||^2$$

and define $\beta : \Re^{n+1} \to \Re_+$ by

$$\beta(z) : = \gamma \min\{1, [\psi(z)]^s\}.$$

Algorithm 4.1

Step 0: Choose constants $\delta \in (0,1)$, $s \in [\frac{1}{2}, 1]$ and $\sigma \in (0, \frac{1}{2})$. Let $t^0 := \bar{t}$, $x^0 \in \Re^n$ be an arbitrary point and $k := 0$.

Step 1: If $H(z^k) = 0$ then stop. Otherwise, let $\beta_k : = \beta(z^k)$.

Step 2: Compute $\Delta z^k := (\Delta t^k, \Delta x^k) \in \Re^n \times \Re^n$ by

$$H(z^k) + H'(z^k)\Delta z^k = \beta_k \bar{z}. \tag{4.3}$$

Step 3: Let l_k be the smallest nonnegative integer l satisfying

$$\psi(z^k + \delta^l \Delta z^k) \leqslant [1 - 2\sigma(1 - \gamma\bar{t})\delta^l]\psi(z^k). \tag{4.4}$$

Define $z^{k+1} := z^k + \delta^{l_k} \Delta z^k$.

Step 4: Replace k by $k+1$ and go to Step 1.

Lemma 4.1. *If for some* $\tilde{z}:=(\tilde{t},\tilde{x}) \in \Re_{++} \times \Re^n$, $H'(\tilde{z})$ *is nonsingular, then there exist a closed neighborhood* $\mathcal{N}(\tilde{z})$ *of* $\tilde{z}$ *and a positive number* $\bar{\alpha} \in (0,1]$ *such that for any* $z=(t,x) \in \mathcal{N}(\tilde{z})$ *and all* $\alpha \in [0,\bar{\alpha}]$ *we have* $t \in \Re_{++}$, $H'(z)$ *is invertible and*

$$\psi(z+\alpha\,\Delta z) \leqslant [1-2\sigma(1-\gamma\bar{t})\alpha]\psi(z), \tag{4.5}$$

where Δz *is the solution of the following linear equation*:

$$H(z)+H'(z)\,\Delta z=\beta(z)\bar{z}.$$

Proof. Since $H'(\tilde{z})$ is invertible and $\tilde{t} \in \Re_{++}$, there exists a closed neighborhood $\mathcal{N}(\tilde{z})$ of $\tilde{z}$ such that for any $z=(t,x) \in \mathcal{N}(\tilde{z})$ we have $t \in \Re^n_{++}$ and that $H'(z)$ is invertible. For any $z \in \mathcal{N}(\tilde{z})$, let $\Delta z=(\Delta t,\Delta x) \in \Re^n \times \Re^n$ be the unique solution of the following equation:

$$H(z)+H'(z)\,\Delta z=\beta(z)\bar{z} \tag{4.6}$$

and for any $\alpha \in [0,1]$, define

$$g_z(\alpha)=G(z+\alpha\,\Delta z)-G(z)-\alpha G'(z)\Delta z.$$

From (4.6), for any $z \in \mathcal{N}(\tilde{z})$,

$$\Delta t=-t+\beta(z)\bar{t}.$$

Then for all $\alpha \in [0,1]$ and all $z \in \mathcal{N}(\tilde{z})$,

$$t+\alpha\,\Delta t=(1-\alpha)t+\alpha\beta(z)\bar{t} \in \Re_{++}. \tag{4.7}$$

It follows from the mean value theorem that

$$g_z(\alpha)=\alpha\int_0^1 [G'(z+\theta\alpha\,\Delta z)-G'(z)]\Delta z d\theta.$$

Since $G'(\cdot)$ is uniformly continuous on $\mathcal{N}(\tilde{z})$ and $\Delta z \to \Delta\tilde{z}$ as $z \to \tilde{z}$, for all $z \in \mathcal{N}(\tilde{z})$,

$$\lim_{\alpha\downarrow 0} ||g_z(\alpha)||/\alpha=0.$$

Then, from (4.7), (4.6) and the fact that $\beta(z) \leqslant \gamma[\psi(z)]^{1/2}$, for all $\alpha \in [0,1]$ and all $z \in \mathcal{N}(\tilde{z})$, we have

$$\begin{aligned}
&(t+\alpha\,\Delta t)^2\\
&\quad=((1-\alpha)t+\alpha\beta(z)\bar{t})^2\\
&\quad=(1-\alpha)^2t^2+2(1-\alpha)\alpha\beta(z)t\bar{t}+\alpha^2[\beta(z)]^2\bar{t}^2\\
&\quad=(1-\alpha)^2t^2+2\alpha\beta(z)t\bar{t}+O(\alpha^2)\\
&\quad\leqslant (1-\alpha)^2t^2+2\alpha\gamma[\psi(z)]^{\frac{1}{2}}||H(z)||\bar{t}+O(\alpha^2)\\
&\quad=(1-\alpha)^2t^2+2\alpha\gamma\bar{t}\psi(z)+O(\alpha^2)
\end{aligned} \tag{4.8}$$

and

$$\begin{aligned}
&\|G(z+\alpha\,\Delta z)\|^2\\
&\quad=\|G(z)+\alpha G'(z)\Delta z+g_z(\alpha)\|^2\\
&\quad=\|(1-\alpha)G(z)+g_z(\alpha)\|^2\\
&\quad=(1-\alpha)^2\|G(z)\|^2+2(1-\alpha)G(z)^{\mathrm{T}}g_z(\alpha)+\|g_z(\alpha)\|^2\\
&\quad=(1-\alpha)^2\|G(z)\|^2+o(\alpha).
\end{aligned}\tag{4.9}$$

It then follows from (4.8) and (4.9) that for all $\alpha\in[0,1]$ and all $z\in\mathcal{N}(\tilde z)$, we have

$$\begin{aligned}
&\psi(z+\alpha\,\Delta z)\\
&\quad=\|H(z+\alpha\,\Delta z)\|^2\\
&\quad=(t+\alpha\,\Delta t)^2+\|G(z+\alpha\,\Delta z)\|^2\\
&\quad\leqslant(1-\alpha)^2t^2+2\alpha\gamma\bar t\psi(z)+(1-\alpha)^2\|G(z)\|^2+o(\alpha)+O(\alpha^2)\\
&\quad=(1-\alpha)^2\psi(z)+2\alpha\gamma\bar t\psi(z)+o(\alpha)\\
&\quad=(1-2\alpha)\psi(z)+2\alpha\gamma\bar t\psi(z)+o(\alpha)\\
&\quad=[1-2(1-\gamma\bar t)\alpha]\psi(z)+o(\alpha)\\
&\quad\leqslant[1-2\rho(1-\gamma\bar t)\alpha]\psi(z)+o(\alpha).
\end{aligned}\tag{4.10}$$

Then, from inequality (4.10), we can find a positive number $\bar\alpha\in(0,1]$ such that for all $\alpha\in[0,\bar\alpha]$ and all $z\in\mathcal{N}(\tilde z)$, (4.5) holds. □

Assumption 4.1. For any $z=(t,x)\in\Re^{n+1}$ with $t>0$, $H'(z)$ is nonsingular.

Lemma 4.2. *Suppose that Assumption* 4.1 *holds. Then Algorithm* 4.1 *is well defined at the kth iteration and for any* $k\geqslant 0$,

$$0<t^{k+1}\leqslant t^k\leqslant\bar t,\tag{4.11}$$

and

$$t^k\geqslant\beta(z^k)\bar t.\tag{4.12}$$

Proof. It follows from Lemma 4.1 that Algorithm 4.1 is well defined at the kth iteration. Now we prove that (4.11) and (4.12) hold by induction. First, $t^0=\bar t>0$. From the design of Algorithm 4.1 and that, for any $z\in\Re^{n+1}$, $\beta(z)\leqslant\gamma<1$, we have

$$t^1=(1-\delta^{l_0})t^0+\delta^{l_0}\beta(z^0)\bar t\leqslant(1-\delta^{l_0})t^0+\delta^{l_0}\gamma\bar t\leqslant t^0$$

and

$$t^0\geqslant\beta(z^0)\bar t.$$

Hence (4.11) and (4.12) hold for $k=0$. Suppose that (4.11) and (4.12) hold for $k\geqslant 0$. We now prove that (4.11) and (4.12) hold for $k+1$. From the design of Algorithm 4.1 we have

$$t^{k+1}=(1-\delta^{l_k})t^k+\delta^{l_k}\beta(z^k)\bar t.$$

Since $t^k \geqslant \beta(z^k)\bar{t}$,

$$t^{k+1} \leqslant (1-\delta^{l_k})t^k + \delta^{l_k}t^k = t^k$$

and

$$t^{k+1} \geqslant (1-\delta^{l_k})\beta(z^k)\bar{t} + \delta^{l_k}\beta(z^k)\bar{t} = \beta(z^k)\bar{t} > 0.$$

So, (4.11) holds for $k+1$. We now prove (4.12) holds for $k+1$ by considering the following two cases.

Case 1: $\psi(z^k) > 1$.

In this case we have $\beta_k = \gamma$. It, therefore, follows from $t^k \geqslant \beta(z^k)\bar{t}$ and $\beta(z) = \gamma\min\{1, [\psi(z)]^s\} \leqslant \gamma$ for any $z \in \mathfrak{R}^{n+1}$ that

$$\begin{aligned}
&t^{k+1} - \beta(z^{k+1})\bar{t}\\
&\quad = t^k + \delta^{l_k}\Delta t^k - \beta(z^k + \delta^{l_k}\Delta z^k)\bar{t}\\
&\quad \geqslant (1-\delta^{l_k})t^k + \delta^{l_k}\beta_k\bar{t} - \gamma\bar{t}\\
&\quad \geqslant (1-\delta^{l_k})\beta_k\bar{t} + \delta^{l_k}\beta_k\bar{t} - \gamma\bar{t}\\
&\quad = (1-\delta^{l_k})\gamma\bar{t} + \delta^{l_k}\gamma\bar{t} - \gamma\bar{t}\\
&\quad = 0.
\end{aligned} \tag{4.13}$$

Case 2: $\psi(z^k) \leqslant 1$.

In this case, we have

$$\psi(z^k + \delta^{l_k}\Delta z^k) \leqslant [1 - 2\sigma(1-\gamma\bar{t})\delta^{l_k}]\psi(z^k) \leqslant 1. \tag{4.14}$$

So

$$\beta(z^k + \delta^{l_k}\Delta z^k) = \gamma[\psi(z^k + \delta^{l_k}\Delta z^k)]^s.$$

Hence, by using the first inequality (4.14), we have

$$\begin{aligned}
&t^{k+1} - \beta(z^{k+1})\bar{t}\\
&\quad = t^k + \delta^{l_k}\Delta t^k - \beta(z^k + \delta^{l_k}\Delta z^k)\bar{t}\\
&\quad = (1-\delta^{l_k})t^k + \delta^{l_k}\beta_k\bar{t} - \gamma[\psi(z^k + \delta^{l_k}\Delta z^k)]^s\bar{t}\\
&\quad \geqslant (1-\delta^{l_k})\beta_k\bar{t} + \delta^{l_k}\beta_k\bar{t} - \gamma[1 - 2\sigma(1-\gamma\bar{t})\delta^{l_k}]^s[\psi(z^k)]^s\bar{t}\\
&\quad = \beta_k\bar{t} - \gamma[1 - 2\sigma(1-\gamma\bar{t})\alpha]^s[\psi(z^k)]^s\bar{t}\\
&\quad = \gamma[\psi(z^k)]^s\bar{t} - \gamma[1 - 2\sigma(1-\gamma\bar{t})\alpha]^s[\psi(z^k)]^s\bar{t}\\
&\quad = \gamma\{1 - [1 - 2\sigma(1-\gamma\bar{t})\alpha]^s\}[\psi(z^k)]^s\bar{t}\\
&\quad \geqslant 0.
\end{aligned} \tag{4.15}$$

So, (4.12) holds for $k+1$. This completes our proof. □

Theorem 4.1. *Suppose that Assumption* 4.1 *holds. Then an infinite sequence* $\{z^k\}$ *is generated by Algorithm* 4.1 *and each accumulation point* $\tilde{z}$ *of* $\{z^k\}$ *is a solution of* $H(z) = 0$.

Proof. It follows from Lemma 4.2 and Assumption 4.1 that an infinite sequence $\{z^k\}$ is generated such that $t^k \geqslant \beta_k \bar{t}$ for all $k \geqslant 0$. From the design of Algorithm 4.1, $\psi(z^{k+1}) < \psi(z^k)$ for all $k \geqslant 0$. Hence the two sequences $\{\psi(z^k)\}$ and $\{\beta(z^k)\}$ are monotonically decreasing. Since $\psi(z^k), \beta(z^k) > 0$ $(k \geqslant 0)$, there exist $\tilde{\psi}, \tilde{\beta} \geqslant 0$ such that $\psi(z^k) \to \tilde{\psi}$ and $\beta(z^k) \to \tilde{\beta}$ as $k \to \infty$. If $\tilde{\psi} = 0$ and $\{z^k\}$ has an accumulation point $\tilde{z}$, then from the continuity of $\psi(\cdot)$ and $\beta(\cdot)$ we obtain $\psi(\tilde{z})=0$ and $\beta(\tilde{z})=0$. Then we obtain the desired result. Suppose that $\tilde{\psi} > 0$ and $\tilde{z}=(\tilde{t},\tilde{x}) \in \mathfrak{R}^n \times \mathfrak{R}^n$ is an accumulation point of $\{z^k\}$. By taking a subsequence if necessary, we may assume that $\{z^k\}$ converges to $\tilde{z}$. It is easy to see that $\tilde{\psi} = \psi(\tilde{z})$, $\beta(\tilde{z}) = \tilde{\beta}$ and $\tilde{t} \geqslant \beta(\tilde{z})\bar{t} > 0$. Then, from Assumption 4.1, $H'(\tilde{z})$ exists and is invertible. Hence, by Lemma 4.1, there exist a closed neighborhood $\mathcal{N}(\tilde{z})$ of $\tilde{z}$ and a positive number $\bar{\alpha} \in (0,1]$ such that for any $z=(t,x) \in \mathcal{N}(\tilde{z})$ and all $\alpha \in [0,\bar{\alpha}]$ we have $t \in \mathfrak{R}_{++}$, $H'(z)$ is invertible and

$$\psi(z + \alpha\,\Delta z) \leqslant [1 - 2\sigma(1 - \gamma\bar{t})\alpha]\psi(z),$$

where Δz is the solution of the following linear equation:

$$H(z) + H'(z)\,\Delta z = \beta(z)\bar{z}.$$

Therefore, for a nonnegative integer l such that $\delta^l \in (0,\bar{\alpha}]$, we have

$$\psi(z^k + \delta^l \Delta z^k) \leqslant [1 - 2\sigma(1 - \gamma\bar{t})\delta^l]\psi(z^k)$$

for all sufficiently large k. Then, for every sufficiently large k, $l_k \leqslant l$ and hence $\delta^{l_k} \geqslant \delta^l$. Then

$$\psi(z^{k+1}) \leqslant [1 - 2\sigma(1 - \gamma\bar{t})\delta^{l_k}]\psi(z^k) \leqslant [1 - 2\sigma(1 - \gamma\bar{t})\delta^l]\psi(z^k)$$

for all sufficiently large k. This contradicts the fact that the sequence $\{\psi(z^k)\}$ converges to $\tilde{\psi} > 0$. So, we complete our proof. □

Assumption 4.2. The level set $L(z^0) = \{z \in \mathfrak{R}^{n+1} |\ \psi(z) \leqslant \psi(z^0)\}$ is bounded.

It follows from Theorem 4.1 that

Corollary 4.1. *Suppose that Assumptions* 4.1 *and* 4.2 *hold. Then an infinite sequence* $\{z^k\}$ *is generated by Algorithm* 4.1 *and there exists at least one accumulation point* $\tilde{z}$ *of* $\{z^k\}$ *such that* $\tilde{z}$ *is a solution of* $H(z) = 0$.

Assumption 4.3. For any $0 < t_1 \leqslant t_2$, the level set

$$L_{[t_1,t_2]}(z^0) = \{x \in \mathfrak{R}^n\colon \psi(t,x) \leqslant \psi(z^0), t \in [t_1,t_2]\}$$

is bounded.

Assumption 4.4. The solution set of Eq. (4.1) is nonempty and bounded.

For any given $t \in \mathfrak{R}$, define $\psi_t(x) : \mathfrak{R}^n \to \mathfrak{R}^+$ by

$$\psi_t(x) = ||G(z)||^2. \tag{4.16}$$

Clearly, $\psi_0(x) = ||\Phi(x)||^2$, and for any fixed $t \in \Re_{++}$, ψ_t is continuously differentiable with the gradient given by

$$\nabla \psi_t(x) = 2(G'_x(z))^T G(z). \tag{4.17}$$

If Assumption 4.1 holds, then $G'_x(z)$ is nonsingular at any point $z = (t, x) \in \Re_{++} \times \Re^n$. For any $z = (t, x) \in \Re \times \Re^n$,

$$\psi(z) = t^2 + \psi_t(x). \tag{4.18}$$

Lemma 4.3 (Facchinei and Kanzow [14]). *Let $C \subset \Re^n$ be a compact set. Then for any $\delta > 0$, there exists a $\bar{t} \in \Re_{++}$ such that*

$$|\psi_t(x) - \psi_0(x)| \leqslant \delta$$

for all $x \in C$ and all $t \in [0, \bar{t}]$.

Theorem 4.2. *Let $f : \Re^n \to \Re$ be continuously differentiable and coercive, i.e.,*

$$\lim_{||x|| \to \infty} f(x) = +\infty.$$

Let $C \subset \Re^n$ be a nonempty and compact set and define m to be the least value of f on the (compact) boundary of C:

$$m := \min_{x \in \partial C} f(x).$$

Assume further that there are two points $a \in C$ and $b \notin C$ such that $f(a) < m$ and $f(b) < m$. Then, there exists a point $c \in \Re^n$ such that $\nabla f(c) = 0$ and $f(c) \geqslant m$.

This theorem is the famous Mountain Pass Theorem, see Theorem 5.3 [14]. We can use it to prove the following theorem.

Theorem 4.3. *Suppose that Assumptions* 4.1, 4.3 *and* 4.4 *hold. Then*

(i) *An infinite sequence $\{z^k = (t^k, x^k)\}$ is generated by Algorithm* 4.1, *and*

$$\lim_{k \to +\infty} H(z^k) = 0, \quad \text{and} \quad \lim_{k \to +\infty} t^k = 0. \tag{4.19}$$

Hence each accumulation point, say $z^* = (0, x^*)$, of $\{z^k\}$ is a solution of $H(z) = 0$.

(ii) *The sequence $\{z^k\}$ is bounded. Hence, there exists at least one accumulation point, say $z^* = (0, x^*)$, of $\{z^k\}$ such that x^* is a solutions of $\Phi(x) = 0$.*

(iii) *If* (4.1) *has an unique solution x^*, then*

$$\lim_{k \to +\infty} x^k = x^*.$$

Proof. (i) It follows from Lemma 4.2 and Assumption 4.1 that an infinite sequence $\{z^k\}$ is generated such that $t^k \geqslant \beta_k \bar{t}$ for all $k \geqslant 0$. From the design of Algorithm 4.1, $\psi(z^{k+1}) < \psi(z^k)$ for all $k \geqslant 0$. Hence the sequences $\{t^k\}$, $\{\psi(z^k)\}$ and $\{\beta(z^k)\}$ are monotonically decreasing. Since

$\psi(z^k), \beta(z^k) > 0$ $(k \geqslant 0)$, there exist $\tilde{\psi}, \tilde{\beta} \geqslant 0$ such that $\psi(z^k) \to \tilde{\psi}$ and $\beta(z^k) \to \tilde{\beta}$ as $k \to \infty$. Suppose that $\tilde{\psi} > 0$. Then, from Lemma 4.2,

$$\lim_{k\to+\infty} t^k = \tilde{t} \geqslant \tilde{\beta}\bar{t} > 0.$$

By Assumption 4.3, the sequence $\{z^k\}$ is bounded. From Theorem 4.1, $\tilde{\psi} = 0$. This contradiction shows that $\tilde{\psi} = 0$, i.e.,

$$\lim_{k\to+\infty} H(z^k) = 0, \quad \text{and} \quad \lim_{k\to+\infty} t^k = 0.$$

(ii) Suppose that the infinite sequence $\{z^k\}$ is not bounded. Then the sequence $\{x^k\}$ is not bounded. Let $\mathscr{S}$ be the solution set of $\Phi(x) = 0$, i.e., the solution set of $\psi_0(x) = 0$. Without loss of generality, assume that $\{||x^k||\} \to \infty$. Hence there exists a compact set $C \subset \Re^n$ with $\mathscr{S} \subset \operatorname{int} C$ and

$$x^k \notin C \tag{4.20}$$

for all k sufficiently large. Let $a \in \mathscr{S}$ is a solution of Eq. (4.1). Then we have

$$\psi_0(a) = 0.$$

Since

$$\bar{m} := \min_{x\in\partial C} \psi_0(x) > 0,$$

we can apply Lemma 4.3 with $\delta := \bar{m}/4$ and conclude that

$$\psi_{t^k}(a) \leqslant \frac{1}{4}\bar{m} \tag{4.21}$$

and

$$m := \min_{x\in\partial C} \psi_{t^k}(x) \geqslant \frac{3}{4}\bar{m} \tag{4.22}$$

for all k sufficiently large. From (i) we have

$$\psi_{t^k}(x^k) \leqslant \frac{1}{4}\bar{m}, \tag{4.23}$$

for all k sufficiently large. Now let us fix an index k such that (4.20)–(4.23) hold. Applying Theorem 4.2 with $b := x^k$, we obtain the existence of a vector $c \in \Re^n$ such that

$$\nabla\psi_{t^k}(c) = 0 \quad \text{and} \quad \psi_{t^k}(c) \geqslant \tfrac{3}{4}\bar{m} > 0.$$

From (4.17) and Assumption 4.1 we have $G(t^k, c) = 0$, i.e., $\psi_{t^k}(c) = 0$, which gives us the desired contradiction.

It follows from (i) and (ii) that (iii) holds. □

Remark. Clearly, if Assumption 4.2 holds, then Assumptions 4.3 and 4.4 hold. But the converse of this statement is not true. For example, let $x = (x_1, x_2) \in \Re^2$, $B = [-1, 1]$, $\Pi_B(\cdot)$ be the projection operator onto B and

$$\Phi(x) = \begin{pmatrix} x_2 \\ x_2 - \Pi_B(x_1 + x_2) \end{pmatrix}.$$

Let

$$H(t,x) = \begin{pmatrix} t \\ x_2 + tx_1 \\ x_2 - (x_1 + x_2)/q(t,x) \end{pmatrix},$$

where

$$q(t,x) = \begin{cases} |t| \ln(\exp(1/|t|) + \exp(\sqrt{(x_1+x_2)^2 + t^2}/|t|)) & \text{if } t \neq 0, \\ \max\{1, |x_1 + x_2|\} & \text{otherwise,} \end{cases}$$

and

$$\psi(t,x) = t^2 + (x_2 + tx_1)^2 + [x_2 - (x_1 + x_2)/q(t,x)]^2.$$

Clearly, $(0,0)$ is the unique solution of $\Phi(x) = 0$, and for any $t^2 \geqslant t^1 > 0$ and $\alpha > 0$, the level set

$$L_{[t^1,t^2]}(\alpha) := \{x \in \Re^2 \colon \psi(t,x) \leqslant \alpha,\ t \in [t^1, t^2]\}$$

is bounded. Let $t^0 = 1$, $x_1^0 = 0$, and $x_2^0 = 1$. Then we have $\psi(t^0, x_1^0, x_2^0) \geqslant 2$. But $\psi(t, x_1, x_2) = 1$ for any $(t, x_1, x_2) \in \Re^3$ with $t = 0$, $x_1 \in [0, +\infty)$ and $x_2 = 1$. So the level set

$$L(z^0) := \{(t, x_1, x_2) \in \Re^3 \colon \psi(t, x_1, x_2) \leqslant \psi(t^0, x_1^0, x_2^0)\}$$

is not bounded.

Let $z^* = (0, x^*)$ and define

$$A(z^*) = \{\lim H'(t^k, x^k) \colon t^k \downarrow 0^+ \text{ and } x^k \to x^*\}. \tag{4.24}$$

Clearly, $A(z^*) \subseteq \partial_B H(z^*)$.

Lemma 4.4. *If all $V \in A(z^*)$ are nonsingular, then there is a neighborhood $N(z^*)$ of z^* and a constant C such that for any $z = (t,x) \in N(z^*)$ with $t \neq 0$, $H'(z)$ is nonsingular and*

$$\|(H'(z))^{-1}\| \leqslant C.$$

Proof. If the conclusion is not true, then there is a sequence $\{z^k = (t^k, x^k)\}$ with all $t^k \neq 0$ such that $z^k \to z^*$, and either all $H'(z^k)$ are singular or $\|(H'(z^k))^{-1}\| \to +\infty$. Since H is locally Lipschitzian, ∂H is bounded in a neighborhood of z^*. By passing to a subsequence, we may assume that $H'(z^k) \to V$. Then V must be singular, a contradiction to the assumption of this lemma. This completes the proof. □

Theorem 4.4. *Suppose that Assumption* 4.1 *is satisfied and z^* is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm* 4.1. *Assume that all $V \in A(z^*)$ are nonsingular and $s \in (\frac{1}{2}, 1]$. If H is semismooth at z^*, then the whole sequence $\{z^k\}$ converges to z^*,*

$$\|z^{k+1} - z^*\| = o(\|z^k - z^*\|) \tag{4.25}$$

and

$$\psi(z^{k+1}) = o(\psi(z^k)). \tag{4.26}$$

Furthermore, if H is strongly semismooth at z^*, then

$$||z^{k+1} - z^*|| = O(||z^k - z^*||^{2s}) \tag{4.27}$$

and

$$\psi(z^{k+1}) = O(\psi(z^k)^{2s}). \tag{4.28}$$

Proof. First, from Theorem 4.1, z^* is a solution of $H(z) = 0$. Then, from Lemma 4.4, there is a scalar C such that for all z^k sufficiently close to z^*,

$$||H'(z^k)^{-1}|| \leqslant C. \tag{4.29}$$

Hence, under the assumption that H is semismooth (strongly semismooth, respectively) at z^*, from Lemma 1.1, for z^k sufficiently close to z^*, we have

$$\begin{aligned} ||z^k + \Delta z^k - z^*|| &= ||z^k + H'(z^k)^{-1}[-H(z^k) + \beta_k \bar{z}] - z^*|| \\ &= O(||H(z^k) - H(z^*) - H'(z^k)(z^k - z^*)|| + \beta_k \bar{t}) \\ &= o(||z^k - z^*||) + O(\psi(z^k)^s) \quad (=O(||z^k - z^*||^2) + O(\psi(z^k)^s)). \end{aligned} \tag{4.30}$$

Then, because H is semismooth at z^*, H is locally Lipschitz continuous near z^*, for all z^k close to z^*,

$$\psi(z^k)^s = ||H(z^k)||^{2s} = O(||z^k - z^*||^{2s}). \tag{4.31}$$

Therefore, from (4.30) and (4.31), if H is semismooth (strongly semismooth, respectively) at z^*, for all z^k sufficiently close to z^*,

$$||z^k + \Delta z^k - z^*|| = o(||z^k - z^*||) \quad (=O(||z^k - z^*||^{2s})). \tag{4.32}$$

By (4.32), for any $\varepsilon \in (0, \frac{1}{2})$, there is a $k(\varepsilon)$ such that for all $k \geqslant k(\varepsilon)$,

$$||z^k + \Delta z^k - z^*|| \leqslant \varepsilon ||z^k - z^*||. \tag{4.33}$$

By (4.3) and (4.29),

$$\begin{aligned} ||\Delta z^k|| &= ||H'(z^k)^{-1}[-H(z^k) + \beta_k \bar{z}]|| \\ &\leqslant C||H(z^k)|| + \bar{t}C[\psi(z^k)]^{1/2} \\ &= (1 + \bar{t})C||H(z^k)||. \end{aligned}$$

Then,

$$\begin{aligned} ||z^k - z^*|| &\leqslant ||\Delta z^k|| + ||z^k + \Delta z^k - z^*|| \\ &\leqslant (1 + \bar{t})C||H(z^k)|| + \varepsilon ||z^k - z^*||. \end{aligned} \tag{4.34}$$

Thus,

$$||z^k - z^*|| \leqslant 2(1 + \bar{t})C||H(z^k)||. \tag{4.35}$$

Hence, if H is semismooth (strongly semismooth, respectively) at z^*, for all z^k sufficiently close to z^*, we have

$$\begin{aligned}\psi(z^k+\Delta z^k) &= ||H(z^k+\Delta z^k)||^2\\ &= O(||z^k+\Delta z^k-z^*||^2)\\ &= o(||z^k-z^*||^2) \quad (=O(||z^k-z^*||^{4s}))\\ &= o(||H(z^k)-H(z^*)||^2) \quad (=O(||H(z^k)-H(z^*)||^{4s}))\\ &= o(\psi(z^k)) \quad (=O(\psi(z^k)^{2s})).\end{aligned} \tag{4.36}$$

Therefore, for all z^k sufficiently close to z^* we have

$$z^{k+1}=z^k+\Delta z^k,$$

which, together with (4.32) and (4.36), proves (4.25) and (4.26), and if H is strongly semismooth at z^*, proves (4.27) and (4.28). So, we complete our proof. □

5. A primal–dual algorithm for minimizing a sum of Euclidean norms

Define $p(t,s)\colon \mathfrak{R}^{d+1}\to\mathfrak{R}^d$ by

$$p(t,s)=\begin{cases}\phi(|t|,s) & \text{if } t\neq 0,\\ \Pi_B(s) & \text{if } t=0,\end{cases} \tag{5.1}$$

where

$$\phi(t,s)=s/q(t,s),\ (t,s)\in\mathfrak{R}_{++}\times\mathfrak{R}^d, \tag{5.2}$$

and

$$q(t,s)=t\ln(\exp(1/t)+\exp(\sqrt{||s||^2+t^2}/t)).$$

Proposition 5.1 (Qi and Zhou [26]). *$p(t,s)$ has the following properties*:

(i) *For any given $t>0$, $p(t,s)$ is continuously differentiable*;

(ii) *$p(t,s)\in\operatorname{int}B$, for any given $t>0$*;

(iii) *$|p(t,s)-\Pi_B(s)|\leqslant(\ln 2+1)t$*;

(iv) *For any given $t>0$,*

$$\nabla p_s(t,s)=\frac{1}{q(t,s)}I_d-\frac{ss^{\mathrm{T}}}{q(t,s)^2(1+\mathrm{e}^{(1-\sqrt{||s||^2+t^2})/t})\sqrt{||s||^2+t^2}}, \tag{5.3}$$

and $\nabla p_s(t,s)$ is symmetric, positive definite and $||\nabla\phi_s(t,s)||<1$;

(v) *For any given $s\in\mathfrak{R}^d$ and $t>0$,*

$$\nabla p_t(t,s)=-\frac{1}{q^2(t,s)}\left(\ln e(t,s)-\frac{\mathrm{e}^{1/t}}{te(t,s)}+\frac{||s||^2\mathrm{e}^{\sqrt{||s||^2+t^2}t/t}}{t\sqrt{||s||^2+t^2}e(t,s)}\right)s, \tag{5.4}$$

where $e(t,s)=\mathrm{e}^{1/t}+\mathrm{e}^{\sqrt{||s||^2+t^2}/t}$;

(vi) *$p(t,s)$ is a strongly semismooth function on $\Re \times \Re^d$;*

(vii) *If $||s^*|| < 1$, then*

$$\lim_{\substack{t^k \downarrow 0^+ \\ s^k \to s^*}} \nabla p_s(t^k, s^k) = I_d;$$

(viii) *If $||s^*|| > 1$, then*

$$\lim_{\substack{t^k \downarrow 0^+ \\ s^k \to s^*}} \nabla p_s(t^k, s^k) = \frac{1}{||s^*||} I_d - \frac{1}{||s^*||^3} s^*(s^*)^{\mathrm{T}},$$

which is symmetric, nonnegative definite, and the norm of this matrix is less than 1 and the rank of this matrix is $d-1$.

Let $z := (t, x, y) \in \Re \times \Re^n \times \Re^{md}$ and define $H : \Re \times \Re^n \times \Re^{md} \to \Re^{n+md+1}$ by

$$H(z) := \begin{pmatrix} t \\ Ap(t,y) - tx \\ y - p(t,y) - (b - A^{\mathrm{T}}x) \end{pmatrix}, \tag{5.5}$$

where $p(t,y) = [p(t,y_1)^{\mathrm{T}}, \ldots, p(t,y_m)^{\mathrm{T}}]^{\mathrm{T}}$. Then H is an augmented smoothing function of F defined in (2.7). From Proposition 5.1, H is continuously differentiable on $\Re_{++} \times \Re^n \times \Re^{md}$ and strongly semismooth on $\Re \times \Re^n \times \Re^{md}$.

Lemma 5.1. *For any $z = (t, x, y) \in \Re_{++} \times \Re^n \times \Re^{md}$,*

$$H'(z) := \begin{pmatrix} 1 & 0 & 0 \\ AE(z) - x & -tI_n & AP(z) \\ -E(z) & A^{\mathrm{T}} & I_{md} - P(z) \end{pmatrix}, \tag{5.6}$$

where

$$E(z) = \nabla p_t(t, y), \tag{5.7}$$

and

$$P(z) = \mathrm{Diag}(p_s'(t, y_i),\ i = 1, \ldots, m), \tag{5.8}$$

and $H'(z)$ is nonsingular.

Proof. We have that (5.6) holds by simple computation. For any $z = (t, x, y) \in \Re_{++} \times \Re^n \times \Re^{md}$, in order to prove $H'(z)$ is nonsingular, we only need to prove that

$$M = \begin{pmatrix} -tI_n & AP(z) \\ A^{\mathrm{T}} & I_{md} - P(z) \end{pmatrix}$$

is nonsingular. For any $t > 0$ and $(x, y) \in \Re^n \times \Re^{md}$, from Proposition 5.1 $P(z)$ is symmetric positive definite and $||P(z)|| < 1$. Let $Mg = 0$, where $g = (g_1^{\mathrm{T}}, g_2^{\mathrm{T}})^{\mathrm{T}} \in \Re^n \times \Re^{md}$. Then we have

$$-tI_n g_1 + AP(z) g_2 = 0, \tag{5.9}$$

and

$$A^{\mathrm{T}}g_1 + (I_{md} - P(z))g_2 = 0. \tag{5.10}$$

From (5.10) we have

$$g_2 = -(I_{md} - P(z))^{-1}A^{\mathrm{T}}g_1. \tag{5.11}$$

Then,

$$(tI_n + AP(z)(I_{md} - P(z))^{-1}A^{\mathrm{T}})g_1 = 0. \tag{5.12}$$

Let

$$B(z) = tI_n + A(I_{md} - P(z))^{-1}P(z)A^{\mathrm{T}}. \tag{5.13}$$

Then $B(z)$ is an n-by-n symmetric positive definite matrix because A has full rank. So $g_1 = 0$. Thus $g = 0$. This indicates that M is nonsingular. So $H'(z)$ is nonsingular. This completes the proof. □

Choose $\bar{t} \in \Re_{++}$ and $\gamma \in (0,1)$ such that $\gamma\bar{t} < 1$. Let $\bar{z} := (\bar{t}, 0, 0) \in \Re \times \Re^n \times \Re^{md}$. Define ψ: $\Re \times \Re^n \times \Re^{md} \to \Re_+$ by $\psi(z) := ||H(z)||^2$, and $\beta : \Re_+ \times \Re^n \times \Re^{md} \to \Re_+$ by $\beta(z) := \gamma \min\{1, \psi(z)\}$.

Algorithm 5.1

Step 0: Choose constants $\delta \in (0,1)$, $\sigma \in (0,1/2)$ and $\varepsilon_1, \varepsilon_2 > 0$ as termination tolerances. Let $z^0 := (\bar{t}, x^0, y^0) \in \Re_{++} \times \Re^n \times \Re^{md}$ and $k := 0$.

Step 1: If $||Ap(t^k, y^k)|| \leqslant \varepsilon_1$ and $|\sum_{i=1}^m ||b_i - A_i^{\mathrm{T}}x^k|| - b^{\mathrm{T}}p(t^k, y^k)| \leqslant \varepsilon_2$ then stop. Otherwise, let $\beta_k := \beta(z^k)$.

Step 2: Compute $\Delta z^k := (\Delta t^k, \Delta x^k, \Delta y^k) \in \Re \times \Re^n \times \Re^{md}$ by

$$H(z^k) + H'(z^k)\Delta z^k = \beta_k \bar{z}. \tag{5.14}$$

Step 3: Let j_k be the smallest nonnegative integer j satisfying

$$\psi(z^k + \delta^j \Delta z^k) \leqslant [1 - 2\sigma(1 - \gamma\bar{t})\delta^j]\psi(z^k). \tag{5.15}$$

Define $z^{k+1} := z^k + \delta^{j_k} \Delta z^k$.

Step 4: Replace k by $k+1$ and go to Step 1.

Lemma 5.2. *Assume that A has rank n. Then the set*

$$\mathscr{S} = \{(x, y) \in \Re^n \times \Re^{md} : F(x, y) = 0\}$$

where F is defined in (2.7), *is nonempty and bounded.*

Proof. From Lemma 2.1 and Lemma 2.1 of Ref. [26], this lemma holds. □

Lemma 5.3. *For any $0 < t_1 \leqslant t_2$ and $\alpha > 0$, the level set*

$$L_{[t_1,t_2]}(\alpha) = \{(x, y) \in \Re^n \times \Re^{md} : \psi(t, x, y) \leqslant \alpha, t \in [t_1, t_2]\}$$

is bounded.

Proof. For any $(x, y) \in L_{[t_1,t_2]}(\alpha)$,

$$\psi(t,x,y) = t^2 + (Ap(t,y) - tx)^2 + \sum_{i=1}^{m}(y_i - p(t,y_i) - (b_i - A_i^{\mathrm{T}}x))^2 \leqslant \alpha.$$

So

$$\sum_{i=1}^{m}(y_i - p(t,y_i) - (b_i - A_i^{\mathrm{T}}x))^2 \leqslant \alpha, \tag{5.16}$$

and

$$(Ap(t,y) - tx)^2 \leqslant \alpha. \tag{5.17}$$

From (5.17) x is bounded. It follows from (5.16) y is bounded. Hence $L_t(\alpha)$ is bounded. □

By Lemmas 5.1, 5.2 and 5.3 and Theorem 4.3, we have

Theorem 5.1. (i) *An infinite sequence* $\{z^k\} \subseteq \Re \times \Re^n \times \Re^{md}$ *is generated by Algorithm* 5.1, *and*

$$\lim_{k\to+\infty} H(z^k) = 0 \quad \textit{and} \quad \lim_{k\to+\infty} t^k = 0. \tag{5.18}$$

Hence each accumulation point, say $z^* = (0,x^*,y^*)$, *of* $\{z^k\}$ *is a solution of* $H(z) = 0$, *and* x^* *and* $P(y^*)$ *are optimal solutions to problems* (1.1) *and* (2.1), *respectively*;

(ii) *The sequence* $\{z^k\}$ *is bounded. Hence there exists at least an accumulation point, say* $z^* = (0,x^*,y^*)$, *of* $\{z^k\}$ *such that* x^* *and* $P(y^*)$ *are optimal solutions to problems* (1.1) *and* (2.1), *respectively*.

(iii) *If problem* (1.1) *has a unique solution* x^*, *then*

$$\lim_{k\to+\infty} x^k = x^*.$$

Let $z^* = (0,x^*,y^*)$ and define

$$A(z^*) = \{\lim H'(t^k,x^k,y^k)\colon t^k \downarrow 0^+, x^k \to x^* \text{ and } y^k \to y^*\}. \tag{5.19}$$

Proposition 5.2. *Let* $M_0(x^*) = \{i : ||b_i - A_i^{\mathrm{T}}x^*|| = 0, i = 1,\dots,m\}$. *If* $\bar{A} = [A_i, i \in M_0(x^*)]$ *is an* $n \times n$ *nonsingular matrix and* $||y_i^*|| < 1$ *for* $i \in M_0(x^*)$, *then all* $V \in A(z^*)$ *are nonsingular.*

Proof. Without loss of generality, we suppose that $||b_i - A_i^{\mathrm{T}}x^*|| = 0$ for $i = 1,\dots,j$ and $||b_i - A_i^{\mathrm{T}}x^*|| > 0$ for $i = j+1,\dots,m$. Then $||y_i^*|| < 1$ for $i = 1,\dots,j$ and $||y_i^*|| > 1$ for $i = j+1,\dots,m$. From Proposition (5.1) and (5.19) we have that for any $V \in A(z^*)$, there exists a sequence $\{z^k = (t^k,x^k,y^k)\}$ such that

$$V = \begin{pmatrix} 1 & 0 & 0 \\ AE^* - x^* & 0 & AP^* \\ -E^* & A^{\mathrm{T}} & I_{md} - P^* \end{pmatrix},$$

where

$$E^* = [E_1^*, \ldots, E_m^*]^{\mathrm{T}},$$

$$(E_i^*)^{\mathrm{T}} = \lim_{\substack{t^k \downarrow 0^+ \\ x^k \to x^* \\ y_i^k \to y_i^*}} \nabla p_t(t^k, y_i^k), \quad \text{for } i = 1, \ldots, m,$$

and

$$P^* = \mathrm{Diag}(P_i^*),$$

$$P_i^* = I_d, \quad \text{for } i = 1, \ldots, j,$$

$$P_i^* = \frac{1}{||y_i^*||} I_d - \frac{1}{||y_i^*||^3} y_i^* (y_i^*)^{\mathrm{T}}, \quad \text{for } i = j+1, \ldots, m.$$

Let

$$M = \begin{pmatrix} 0 & AP^* \\ A^{\mathrm{T}} & I_{md} - P^* \end{pmatrix}.$$

Hence, proving V is nonsingular is equivalent to proving M is nonsingular.

Let

$$\tilde{A} = [A_{j+1}, \ldots, A_m],$$
$$D = \mathrm{Diag}(P_i^*, i = j+1, \ldots, m),$$

and

$$q = [q_1^{\mathrm{T}}, q_2^{\mathrm{T}}, q_3^{\mathrm{T}}]^{\mathrm{T}} \in \mathscr{R}^n \times \mathscr{R}^n \times \mathscr{R}^{md-n}.$$

Let $Mq = 0$. Then we have

$$\bar{A} q_2 + \tilde{A} D q_3 = 0, \tag{5.20}$$

$$\bar{A}^T q_1 = 0, \tag{5.21}$$

and

$$\tilde{A}^{\mathrm{T}} q_1 + (I_{md-n} - D) q_3 = 0. \tag{5.22}$$

From (5.21) we have $q_1 = 0$. Then, from (5.22) $q_3 = 0$. It follows from (5.20) $q_2 = 0$. Thus $q = 0$. This indicates that M is nonsingular. So V is nonsingular. This completes the proof. □

From Proposition 5.2 and Theorem 4.4 we have

Theorem 5.2. *Suppose that $z^* = (0, x^*, y^*)$ is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm* 5.1. *Let $M_0(x^*) = \{i\colon ||b_i - A_i^{\mathrm{T}} x^*|| = 0, i = 1, \ldots, m\}$.* If *$\bar{A} = [A_i, i \in M_0(x^*)]$ is an $n \times n$ nonsingular matrix and $||y_i^*|| < 1$ for $i \in M_0(x^*)$, then the whole sequence $\{z^k\}$ converges to z^*, and the convergence is Q-quadratic.*

Table 1
Numerical results for Algorithm 5.1

Example	n	d	m	It	NH	NO	$f(x^k)$	relgap	$\|\|Ap\|\|$
1	16	2	17	17	54	4	2.54e + 01	2.70e − 16	7.66e − 16
2	4	2	5	10	17	1	4.00e + 02	0	1.40e − 16
3	10	2	55	14	29	2	2.26e + 02	1.25e − 16	2.41e − 15
4	3	3	100	7	25	0	5.59e + 02	1.02e − 15	1.41e − 13
5	4	4	150	8	25	0	8.46e + 02	6.71e − 16	1.84e − 13
6	5	5	200	7	23	0	1.32e + 03	2.07e − 15	3.69e − 13
7	7	7	300	8	23	0	2.32e + 03	1.96e − 16	1.22e − 13
8	8	8	400	7	21	0	3.48e + 03	1.83e − 15	5.62e − 13
9	9	9	500	7	21	0	4.58e + 03	1.59e − 15	4.55e − 13
10	10	2	100	18	36	2	6.86e + 01	0	9.57e − 14
11	20	3	200	32	104	0	1.78e + 02	3.17e − 16	6.27e − 13

6. Numerical experiments

Algorithm 5.1 was implemented in MATLAB and was run on a DEC Alpha Server 8200 for the following examples, where Examples 1 and 2 are taken from [29] and Examples 3–11 from [18]. Throughout the computational experiments, we used the following parameters:

$$\delta = 0.5, \quad \sigma = 0.0005, \quad \bar{t} = 0.002, \text{ and } \gamma = 0.5.$$

We terminated our iteration when one of the following conditions was satisfied

(1) $k > 50$;
(2) $\text{relgap}(x^k, p(t^k, y^k)) \leqslant$ 1e−8 and $||Ap(t^k, y^k)|| \leqslant$ 1e−12;
(3) ls > 20.

where ls was the number of line search at each step and

$$\text{relgap}(x, p(t, y)) = \frac{\left|\sum_{i=1}^m ||b_i - A_i^T x|| - b^T p(t, y)\right|}{\sum_{i=1}^m ||b_i - A_i^T x|| + 1}.$$

The numerical results which we obtained are summarized in Table 1. In this table, n, d and m specify the problem dimensions, It denotes the number of iterations, which is also equal to the number of Jacobian evaluations for the function H, NH denotes the number of function evaluations for the function H, NO indicates the number of norms that are zero at the optimal solution, more precisely, which is interpreted as being zero if it is less than the tolerance 10^{-10}, $f(x^k)$ denotes the value of $f(x)$ at the final iteration, relgap denotes the relative duality gap, and $||Ap||$ denotes the value of $||Ap(t, y)||$ at the final iteration. In the following, we give a brief description of the tested problems, where $\mathbf{0}$ is the vector of all zeros and $\mathbf{e}$ is the vector of all ones.

Example 1. This is a SMT problem [29]. The starting point $x^0 = \mathbf{e}$ and $y^0 = \mathbf{0}$.

Example 2. This is a SMT problem [29]. The starting point $x^0 = \mathbf{e}$ and $y^0 = \mathbf{e}$.

Example 3. This is a multifacility location problem [18]. The starting point $x^0 = \mathbf{e}$ and $y^0 = \mathbf{0}$.

The following examples are generated randomly. We use the following pseudo-random sequence:

$$\psi_0 = 7, \quad \psi_{i+1} = (445\psi_i + 1) \bmod 4096, \quad i = 1, 2, \ldots,$$

$$\bar{\psi}_i = \frac{\psi_i}{4096}, \quad i = 1, 2, \ldots .$$

Example 4 (see Overton [18]).

$$n = 3, \ d = 3, \ m = 100.$$

$$A_i = I, \ i = 1, 2, \ldots, m, \ \text{except } A_i = 100I \text{ if } i \bmod 10 = 1.$$

The elements of b_i, $i = 1, 2, \ldots, m$, are successively set to be $\bar{\psi}_1, \bar{\psi}_2, \ldots,$ in the order $(b_1)_1, \ldots, (b_1)_d$, $(b_2)_1, \ldots, (b_m)_d$, except that the appropriate random number is multiplied by 100 to give $(b_i)_j$ if $i \bmod 10 = 1$. The starting point $x^0 = b_m$ and $y^0 = \mathbf{0}$.

Example 5. The same as Example 4 except $n = 4, d = 4, \ m = 150$.

Example 6. The same as Example 4 except $n = 5, d = 5, \ m = 200$.

Example 7. The same as Example 4 except $n = 7, \ d = 7, \ m = 300$.

Example 8. The same as Example 4 except $n = 8, \ d = 8, \ m = 400$.

Example 9. The same as Example 4 except $n = 9, \ d = 9, \ m = 500$.

Example 10 (see Overton [18]).

$$n = 10, \ d = 2, \ m = 100.$$

The elements of $A_i, i = 1, 2, \ldots, m$, those of $b_i, i = 1, 2, \ldots, m$, and those of x^0 are successively set to $\bar{\psi}_1, \bar{\psi}_2, \ldots,$ in the order:

$$(A_1)_{11}, (A_1)_{21}, \ldots, (A_1)_{n1}, (A_1)_{12}, \ldots, (A_1)_{nd}, \ldots, (A_m)_{nd},$$

$$(b_1)_1, \ldots, (b_1)_d, (b_2)_1, \ldots, (b_m)_d, x_1^0, \ldots, x_n^0,$$

except that the appropriate random number is multiplied by 100 to give $(A_i)_{jk}$ or $(b_i)_j$ if $i \bmod 10 = 1$, and $y^0 = \mathbf{0}$.

Example 11. The same as Example 10 except $n = 20, \ d = 3, \ m = 200$.

The results reported in Table 1 show that this method is extremely promising. The algorithm was able to solve all examples after a small number of iterations. The number of iterations increases slowly with the size of the problem. For Examples 1 and 2, the number of iterations required by our algorithm is fewer than that required by the algorithm proposed in [29].

7. Conclusions

In this paper we first transformed the problem of minimizing a sum of norms and its dual problem into a system of strongly semismooth equations, and gave some uniqueness theorems for this problem. We then presented a primal–dual algorithm for this problem by solving this system of strongly semismooth equations. Numerical results showed that this primal–dual algorithm worked very satisfactorily for the tested problems.

Acknowledgements

The authors would like to thank the referee for the helpful comments.

References

[1] J. Alpert, T.F. Chan, A.B. Kahng, I.L. Markov, P. Mulet, Faster minimization of linear wirelength for global placement, IEEE Trans. Comput. Aided Design Integrated Circuits and Systems 17 (1998) 3–13.
[2] K.D. Andersen, An efficient Newton barrier method for minimizing a sum of Euclidean norms, SIAM J. Optim. 6 (1996) 74–95.
[3] K.D. Andersen, E. Christiansen, Minimizing a sum of norms subject to linear equality constraints, Comput. Optim. Appl. 11 (1998) 65–79.
[4] K.D. Andersen, E. Christiansen, A.R. Conn, M.L. Overton, An efficient primal–dual interior-point method for minimizing a sum of Euclidean norms, SIAM J. Sci. Comput. 22 (2000) 243–262.
[5] P.H. Calamai, A.R. Conn, A stable algorithm for solving the multifacility location problem involving Euclidean distances, SIAM J. Sci. Statist. Comput. 1 (1980) 512–526.
[6] B. Chen, P.T. Harker, A continuation method for monotone variational inequalities, Math. Programming 69 (1995) 237–253.
[7] B. Chen, P.T. Harker, Smooth approximations to nonlinear complementarity problems, SIAM J. Optim. 7 (1997) 403–420.
[8] C. Chen, O.L. Mangasarian, Smoothing methods for convex inequalities and linear complementarity problems, Math. Programming 71 (1995) 51–69.
[9] C. Chen, O.L. Mangasarian, A class of smoothing functions for nonlinear and mixed complementarity problems, Comput. Optim. Appl. 5 (1996) 97–138.
[10] X. Chen, L. Qi, D. Sun, Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities, Math. Comput. 67 (1998) 519–540.
[11] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[12] T. De Luca, F. Facchinei, C. Kanzow, A semismooth equation approach to the solution of nonlinear complementarity problems, Math. Programming 75 (1996) 407–439.
[13] F. Facchinei, C. Kanzow, A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems, Math. Programming 76 (1997) 493–512.
[14] F. Facchinei, C. Kanzow, Beyond monotonicity in regularization methods for nonlinear complementarity problems, SIAM J. Control Optim. 37 (1999) 1150–1161.

[15] A. Fischer, Solution of monotone complementarity problems with locally Lipschitzian functions, Math. Programming 76 (1997) 513–532.
[16] H. Jiang, Smoothed Fischer–Burmeister equation methods for the nonlinear complementarity problem, Preprint, Department of Mathematics, the University of Melbourne, Vic. 3052, Australia, June 1997.
[17] M.R. Osborne, Finite Algorithm in Optimization and Data Analysis, Wiley, New York, 1985.
[18] M.L. Overton, A quadratically convergent method for minimizing a sum of Euclidean norms, Math. Programming 27 (1983) 34–63.
[19] H.-D. Qi, A regularized smoothing Newton method for box constrained variational inequality problems with P_0 functions, SIAM J. Optim. 10 (2000) 251–264.
[20] H.-D. Qi, L.Z. Liao, A smoothing Newton method for extended vertical linear complementarity problems, SIAM J. Matrix Anal. Appl. 21 (2000) 45–66.
[21] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, Math. Oper. Res. 18 (1993) 227–244.
[22] L. Qi, On superlinear convergence of quasi-Newton methods for nonsmooth equations, Oper. Res. Lett. 20 (1997) 223–228.
[23] L. Qi, X. Chen, A globally convergent successive approximation method for severely nonsmooth equations, SIAM J. Control Optim. 33 (1995) 402–418.
[24] L. Qi, D. Sun, G. Zhou, A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities, Math. Programming 87 (2000) 1–35.
[25] L. Qi, J. Sun, A nonsmooth version of Newton's method, Math. Programming 58 (1993) 353–367.
[26] L. Qi, G. Zhou, A smoothing Newton method for minimizing a sum of Euclidean norms, SIAM J. Optim. 11 (2000) 389–410.
[27] D. Sun, A regularization Newton method for solving nonlinear complementarity problems, Appl. Math. Optim. 40 (1999) 315–339.
[28] E. Weiszfeld, Sur le point par lequel la somme des distances de n points donnes est minimum, Tohoku Math. J. 43 (1937) 355–386.
[29] G. Xue, Y. Ye, An efficient algorithm for minimizing a sum of Euclidean norms with applications, SIAM J. Optim. 7 (1997) 1017–1036.
[30] G. Xue, Y. Ye, An efficient algorithm for minimizing a sum of P-norms, SIAM J. Optim. 10 (2000) 551–579.
[31] N. Yamashita, M. Fukushima, Modified Newton methods for solving semismooth reformulations of monotone complementarity problems, Math. Programming 76 (1997) 273–284.
[32] G. Zhou, D. Sun, L. Qi, Numerical experiments for a class of squared smoothing Newton methods for box constrained variational inequality problems, in: M. Fukushima, L. Qi (Eds.), Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Kluwer Academic Publishers, Boston, 1999, pp. 421–441.